

# Deriving Insights from National Happiness Indices

Anthony Brew, Derek Greene, **Daniel Archambault**, and  
Pádraig Cunningham

Clique Strategic Research Cluster, UCD Dublin, Ireland

# Rating: PG-13

Warning: These visualizations contain coarse language.  
Viewer discretion is advised.

# Outline

- 1 Related Work
- 2 Macro-Analysis
- 3 Micro-Analysis and SentireCrowds
- 4 Conclusion and Future Work

# Twitter as a Data Source

- Twitter
  - Microblogging service - share content via short text updates
  - By 2011, over 200 million users and 200 million posts per day
- Can track content by non-reciprocal follower information
- Data is mostly text and link information
- Opinion and sentiment rich information

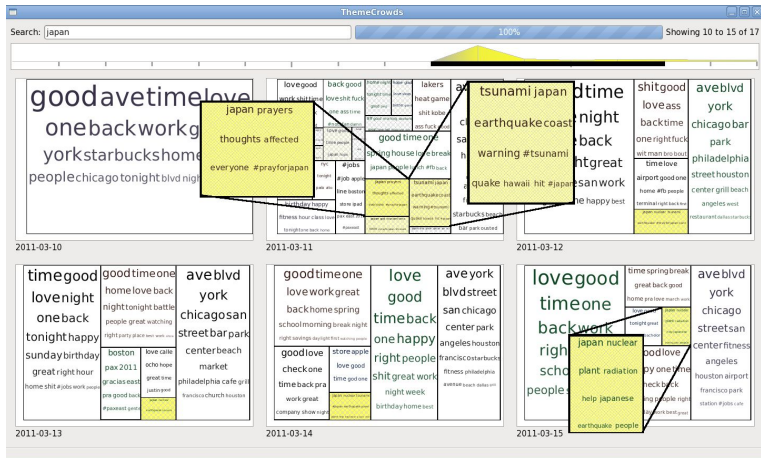
# Motivation

- Industry and academics are interested in sentiment
  - What do people think about a government policy?
  - How do they feel about a product?
  - Can be done at a macroscopic level to see overall sentiment
- We also need to understand micro-level sentiment
  - What aspects of the policy are positive or negative?
  - Do people really like one feature and not another?
- Difficult because unstructured text format of Twitter data
  - noise and errors in the data
  - restricted to 140 characters
- Goals:
  - Macroscopic presentation of mined sentiment
  - Visualization of context of that sentiment

# Twitter Data Analysis

- Community structure and growth
  - communities and interests (Java *et al.*, 2007)
  - influential users and topic spread (Kwak *et al.*, 2010)
  - topics during live event (Shamma *et al.*, 2009)
- Opinion Mining
  - Classifier for positive-negative tweets (Pak *et al.*, 2010)
  - Noisy data classification (Davidov *et al.*, 2010)
- Visualization of trending topics
  - visualization of conversations (Dörk *et al.*, 2010)
  - streamgraphs for tag clouds (Shi *et al.*, 2010)

# ThemeCrowds: Twitter Topic Micro-Analysis



- System to track what people say about topics over time

# Data Set

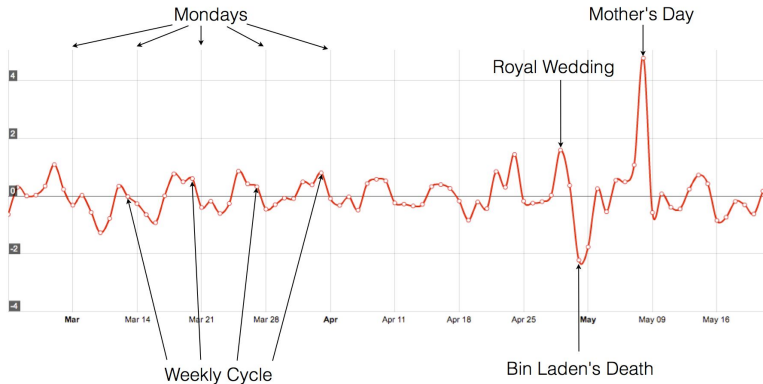
- Twitter corpus collected from nine US cities
- Collected from March 1st 2011 through May 21st 2011
- About 13 million tweets and 337 thousand users



# Macro-Analysis Concept

- Express sentiment as an index
- Use Facebook happiness index (Kramer, 2010)
  - $H_d = \frac{\mu_{pd} - \mu_p}{\sigma_p} - \frac{\mu_{nd} - \mu_n}{\sigma_n}$
  - $\mu_{id}$  - positive/negative terms on a given day
  - $\mu_i$  - overall daily averages
  - $\sigma_i$  - standard deviation over all days analyzed
- Difference of percentage of positive to negative
- Scaled according to data set norms
- 507 positive and 603 negative terms drawn from the Dictionary of Affect in Language

# Macro-Analysis line graph



- Weekly cycle detected with happiness maximized on weekends
  - confirms Kramer's findings on Facebook data
- Royal Wedding, Bin Laden's Death, and Mother's Day indicated

# Terms of Increased Usage

Event	Raw Frequency	Increased Usage
<i>Royal Wedding</i>	lol st like not good love th ave #jobs go	#royalwedding wedding royal #ff st kate ave prince friday york
<i>Bin Laden's Death</i>	lol like not osama st laden good go dead love	osama laden dead obama usa news killed death us president
<i>Mother's Day</i>	lol mothers happy like not love st :) good go	mothers happy mom lakers #if- oumarryme moms love mother #hap- pymothersday :)

- Increased usage compared to previous seven days as baseline

# Bigram Analysis

Event	Sentiment-Associated Terms	Term-Sentiment Bi-grams
<i>Royal Wedding</i>	royal #royalwedding #ff kate prince william dress friday watching #icant- standpeoplethat	royal-wedding watching- wedding watch-wedding #royalwedding-wedding #royalwedding-like kate- wedding friday-happy fri- accident #royalwedding- not royal-not
<i>Bin Laden's Death</i>	osama laden obama news president america usa sunday right #ileftyoube- cause	laden-dead osama-dead osama-killed laden- killed laden-us osama-us obama-dead osama-death god-bless osama-not
<i>Mother's Day</i>	mom lakers #ifyoumar- ryme #happymothers- day #factsaboutmymom brunch shes mavs church kobe	mom-mothers mom-happy mom-love brunch-mothers dinner-mothers world- mothers one-mothers mom-:) world-happy family-mothers

- Co-occurring terms with words in sentiment lexicon

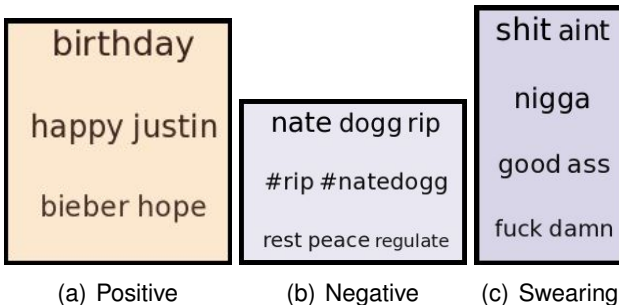
# Micro-analysis

- Macro-analysis reveals coherent sentiment topics
- SentireCrowds
  - What are groups of users saying about a given topic?
  - Are there multiple groups of users?
- SentireCrowds summarizes what groups of users are saying over time.

# Approach

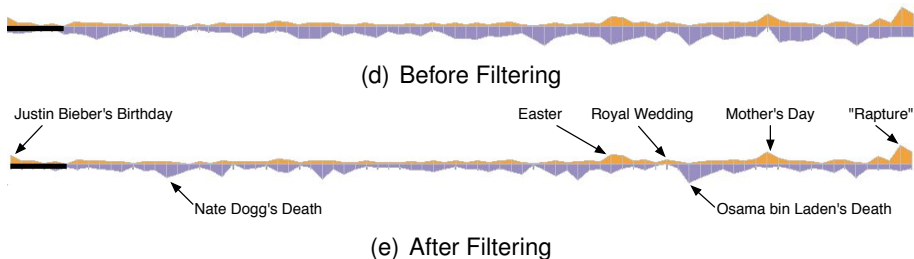
- ThemeCrowds Approach:
  - divide data into 24 hour snapshots
  - create a profile document per user on that day
  - hierarchically cluster documents via textual similarity
  - visualize with multilevel treemap
- Automatically find antichains of maximal sentiment to display
- Apply macro-analysis score on a per cluster basis

# Positive, Negative, and Swearing Clusters



- Clusters coloured using sentiment scores
  - positive in tan
  - negative in purple
- Highly negative swearing clusters present
- Can be filtered out by relaxed content matching

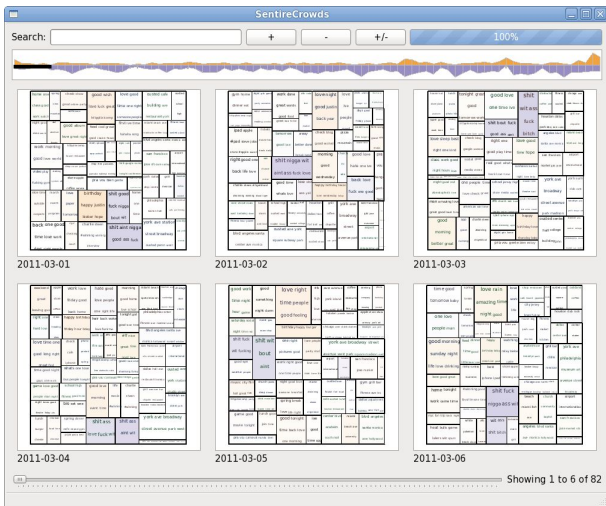
# Timeline



- Explicative clusters tend to create lots of negative sentiment
- After removal, begin to see critical positive and negative events



## Live Demonstration



# Conclusion and Future Work

- System maintains a happiness index for Twitter
  - performed analysis on macroscopic signal
  - SentireCrowds on a groups of users level
- Exploration in terms of variable topic-sentiment bigrams
- User experimentation for effectiveness